Business Intelligence Methods for Sustainable Development of the Railways

Aida-Maria POPA
Academy of Economic Studies, Bucharest, Romania
aida_popa@yahoo.com

This paper aims to present a new approach of business intelligence technologies in the context of sustainable development of the railways. The concept of business intelligence is increasingly used in the developed companies and considering that the current economic market is more dynamic from year to year, business intelligence solutions plays an important role for companies to be able to develop efficient plans for both short-term and medium and long term developing. This paper will focus on two technologies: data-warehouse and data-mining and how are they use in the railway business. The subject adapts to the current development trend of European countries to direct the transport of freight and passengers to the railway for support environment.
Keywords: *sustainable development, business intelligence, data-warehouse, data-mining*

1 Introduction

Railway has demonstrated in the last two centuries an important contribution to the prosperity of states, to the development of national and global economies but also to the possibility of transportation in any point on the planet.
By its nature, the railway is composed of complex processes and activities which are conducted continuously while on a network having significant geographical spread. It involves humans, materials and financial resources with unpredictable events, hence the need for strong systems to respond with flexibility and accuracy to determine market requirements and increase the quality of services offered to all customer`s categories.
The evidence of tradition of Informatics Romanian Railway dates from July 1, 1967 when it was founded the Informatics Centre for Rail Transport (then called Electronic mechanized Computing Center Ministry of Railways). In short time, it has succeeded in establishing itself as a strong core and it managed to maintain and strengthen this status even in the present form of reorganization and transition.
Through activity carried out over the years since it was founded, the Informatics Centre for Rail Transport consistently succeeded to bring beneficial solutions to its customers by deploying applications and advanced information systems. An important aspect of these systems is to lower costs due the fact that they were made in the railway sector, which brought another benefit that is a closely knowledge about processes and activities that were to be computerized.
Currently, the main objective of Romanian Railways is to develop business activity by increasing efficiency and ensuring a future of rail transport, keeping up with modern times and considering the ever present main competitor in the transport field, namely road transport. To achieve this objective, the Romanian Railways is going through a restructuring process that is based on substantial investments.

2. Some aspects regarding Business Intelligence

According to the Romanian Association of Economic Intelligence, business intelligence represents all the activities of research, collecting, processing and information dissemination of useful economic agents in order to gain competitive advantage by exploiting them in the defensive and / or offensive manner.
The paper "Business Intelligence Roadmap" [1] presents business intelligence as an architecture and a collection of applications and operational integrated databases and decision support systems, which provide the

business community easier access to business data.

According to Brândaș C. [2] the most important objectives of Business Intelligence are:

- Gathering and analyzing large volumes of data and information extracted from both the operational database and the data warehouse within the organization;
- The combination of two processes, the knowledge management and the decision management;
- Obtaining complex information for managers and competitive advantage by exploiting technologies to support decision making within the organization.

Website [3] defines Business Intelligence (BI) as a process that aims to serve and support business process management. Streamlining the process through precise timely and correctly founded decisions is the main purpose of the choice of implementing a Business Intelligence system. This process consists people, a set of tools and specialized software applications, methods, services and techniques, all of which been necessary for collecting, storage, analysis and better use of data and information derived from the other processes that occur within the organization. Operation process and also business decision making, depends entirely on the aforementioned elements.

Architectural principles of a business intelligence system are [4]:

• scalability and high performance;
• complete functionality;
• continuous development
• openness and extensibility;
• developing and fast running.

2. Sustainable development

2.1. Definition of the concept of sustainable development

What sustainability means is still no universally agreed definition on. There are a variety of views on what it is and how it can be obtained. The concept of sustainability stems from the idea of sustainable development which has become a usual terminology at the World's first Earth Summit in Rio in 1992 (Brundtland Report for the World Commission on Environment and Development).

The original definition of sustainable development is usually considered to be:

"Sustainable development is development that meets the needs of the present without compromising the ability of future generations to meet their own needs. It contains within it two key concepts:

- the concept of needs, in particular the essential needs of the world's poor, to which overriding priority should be given; and
- the idea of limitations imposed by the state of technology and social organization on the environment's ability to meet present and future needs."[5][6]

The concept of sustainable development can be characterized in many different ways, through the following elements:

- Economics: efficiency, growth, stability;
- Society: standard of living, equity, social dialogue and delegation of responsibilities, the protection of culture / heritage;
- Ecology: conservation and protection of natural resources, biodiversity, avoiding pollution.

Development can be considered sustainable when it meets together economic, social and environment objectives.

Sustainable development is a fundamental objective for the European Union. The purpose for which it addressed this concept is continually improving the quality of life of present and future prosperity, an approach that takes into account the integration of economic development, environment and society.

The purpose for which it addressed this concept is continually the quality of life, improving present and future prosperity, an approach that takes into account the

integration of economic development, environment and society.

2.2. The advantages of rail transport

Sustainable development shows its concern for saving energy, air pollution and numerous traffic accidents. Therefore trends were followed for reinstatement on the main plan of rail transportation, in this way being advancing the development of high-speed rail between cities and the corridors with a high population density.

Many international studies have shown clearly that railway is the best way to transport relating to protect the environment.

In Figure 1 highlights the advantages of rail versus other transport systems against four important aspects: energy consumption, exhaust emissions, traffic safety and the land surface required infrastructure.

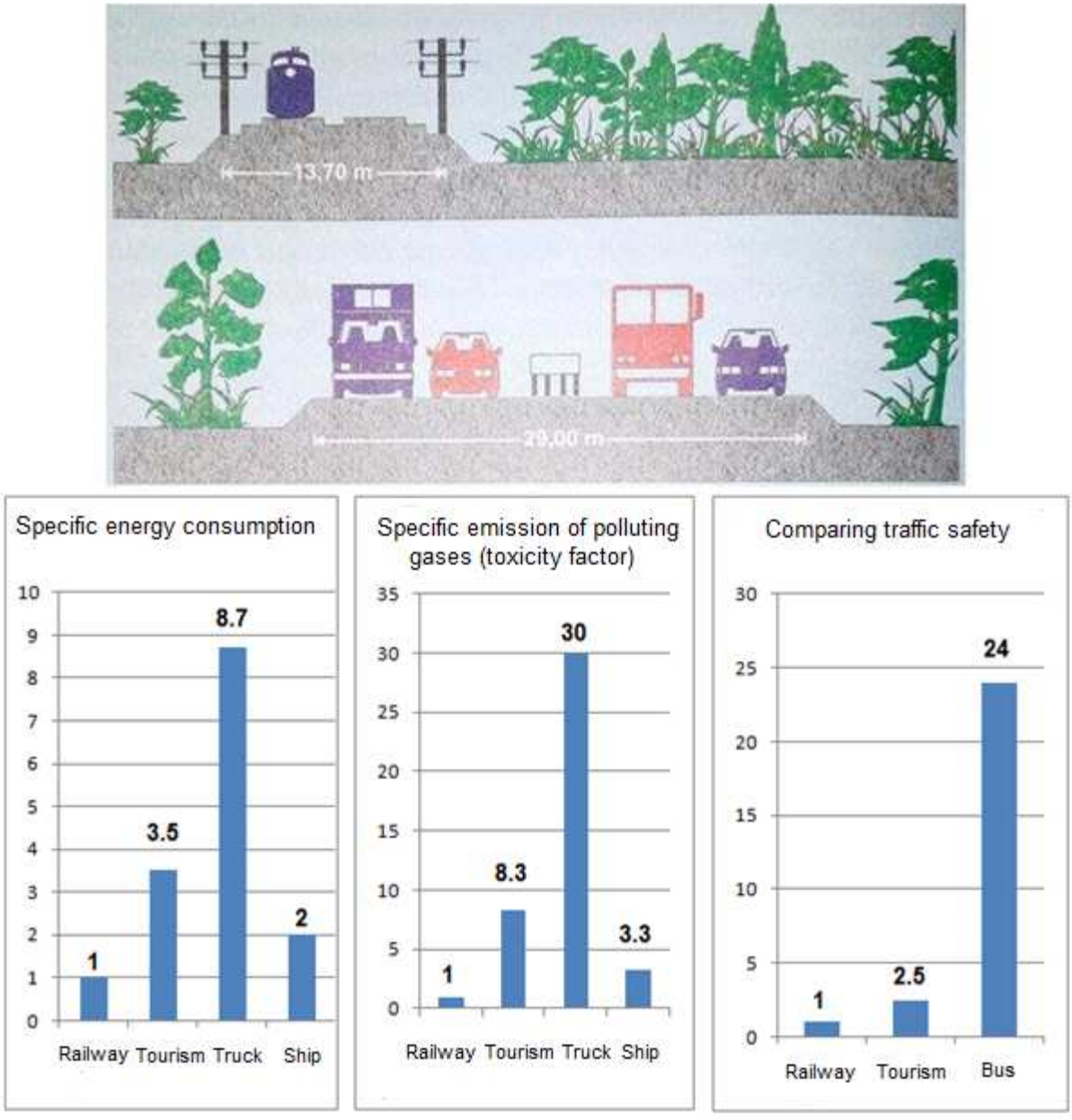

Fig. 1. The advantages of rail transport compared with other transport systems [7]

In terms of environmental pollution we clearly observed advantages of railway transport compared with other systems, the concentration of pollutants is even up to 30 times lower than the concentration of a truck emissions, in the case of freight transport.

By comparing energy consumption registered at various ways of transport, we see a consumption up to 3.5 times higher in case of transport by road and twice as high in case of ship transport compared to rail passenger transport. For freight transport, consumption energy registered by trucks becomes 8.7 times higher than consumption energy registered by rails.

The highest degree of traffic safety is registered by rail and air transport. If we refer to a similar passenger transport capacity, accident risk of the bus is 24 times higher than rail transport. However, in Romania, the risks of road transport is far outweigh of value from the graphic.

Given the fact that Romania belongs in the category of medium populated states, with a population density of about 97 inhabitants per km², land areas are difficult to access for developing transport infrastructure. In case of a modern railway are used about 14 meters wide surface transport compared to the same highway transport capacity that will have 4 lanes and will occupy about 31.5 meters width. Another disadvantage of road transport is also the material that is used to make highway which is not always good for the environment.

3. Data-Mining

According to [8] is presented a methodology called "Methodology for Railway Demand Forecasting Using Data Mining ". Considering the genuine complexity of processes of Knowledge Discovery in Databases (KDD), it was developed this methodology which is based on principles of activity planning. The steps belongs to the knowledge discovery process are set before being executed taking into account the objectives of each KDD request. The application of the methodology is split in four phases as is illustrated by Figure 2.

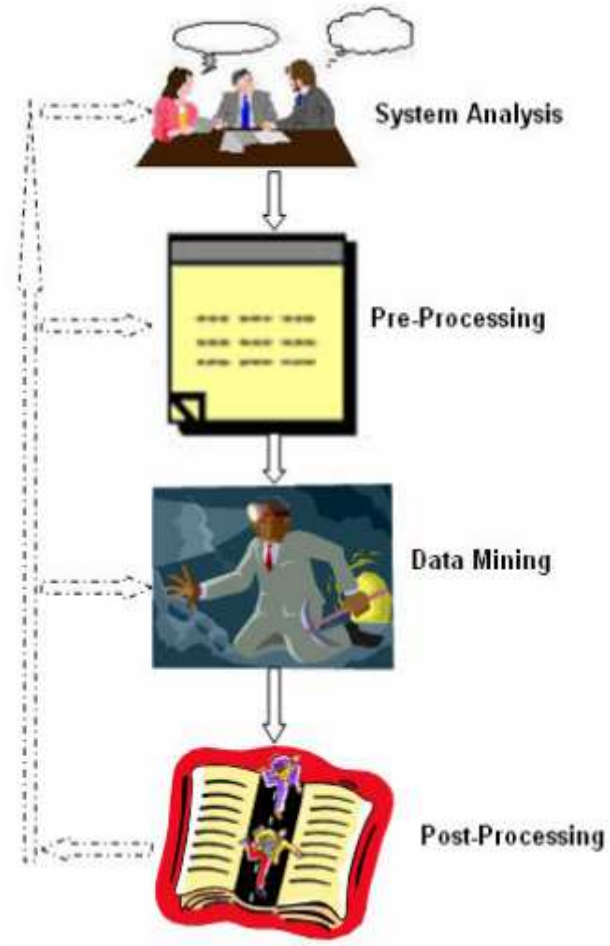

Fig.2. Methodology for Railway Demand Forecasting Using Data Mining [8]

This methodology conjures that is an interactive and iterative process. This translates into the fact that the KDD analysts can return to every step previously browsed if they want to find better results, other than those was already discovered. To achieve this requirement, it is required a precise and detailed documentation about developed actions and achieved results. In this sense it is preferred to use documentation models to have the opportunity to choose which procedures to be adopted taking into account the number and diversity of situations and possibilities.

The phases of methodology are:

a) System Analysis (Figure 3) is the first phase of methodology. The most important

objective of this phase is to define the types of requests that have to be performed by applying techniques of KDD process. The wish in this regard is to identify process objectives and also their resolving or improving. The activities contained in the system analysis phase are: definition of the actors, description of the problem, definition of the objective, expectations and deadline.

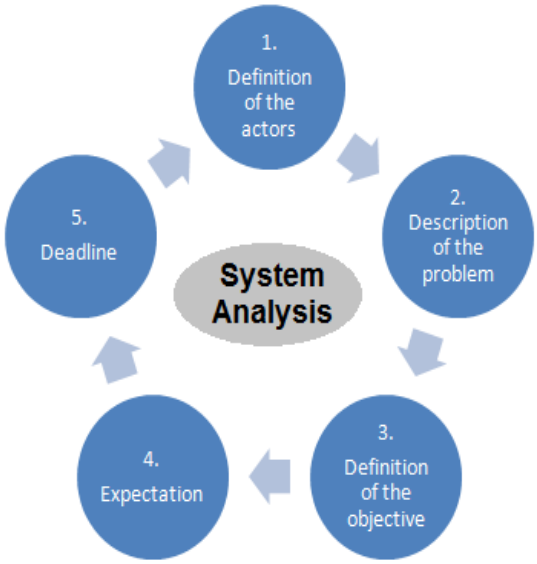

Fig.3. System analysis activities

b) Pre-Processing (Figure 4) is the second phase of methodology. Capturing, organizing, treating and preparing are the functions that make up this phase. These functions are predecessors to Data Mining phase, having a colossal importance for the knowledge discovery process.

After defining the desired outcomes it is going the first activity, choice of the technique, after which we have to choose which techniques can be used that are more close to obtaining results with higher precision.

The activity of selection of the data is indispensable for the pre-processing phase. The selection is necessary for reporting the origin of the information regardless of their source (transactional or from data-warehouse).

Cleaning the data is an optional activity, being used only in case of absent information, inconsistencies, and values that are not pertaining to the domain. According to Kimball and Ross [9] in case of using the information from a data-warehouse the possibility and necessity of cleaning is lower because the creation of a data-warehouse has a process when the database is cleaning.

Codification is the pre-processing activity on which depends the data representation during the KDD process.

The activity of normalization is achieved by assigning a new range to an attribute so that the values could be within the new range in a specified interval (for example, from -1.0 to 1.0 or from 0.0 to 1.0). The adjustments become needful especially for preventing some of the attributes from having a larger range of values than others, for not influence the tendencies of the algorithms of Data Mining that are used.

Enrichment represents the capacity to add additional information to the already recorded data so that these provide more entities to the knowledge discovery process.

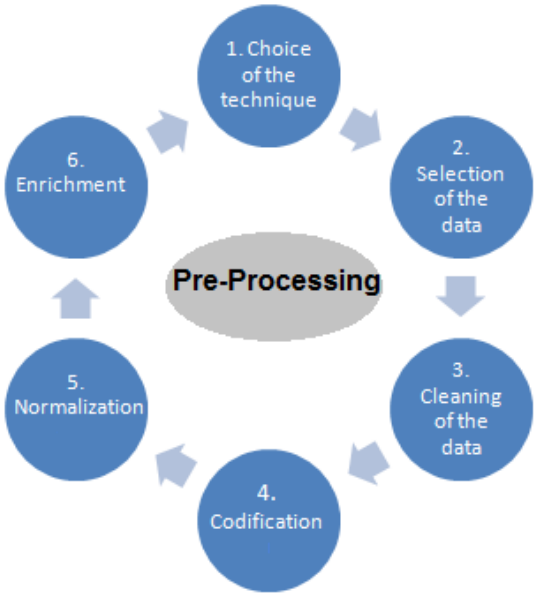

Fig.4. Pre-processing activities

c) Data Mining (Figure 5) is the principal phase of the proposed methodology having the role in the helpful search for new and useful knowledge obtained from the data. This is why the Data Mining and the KDD process are referred by many authors as

two terms that are synonymous. The activities contained in the Data Mining phase are: partition of the data, choice of the tool and data mining.

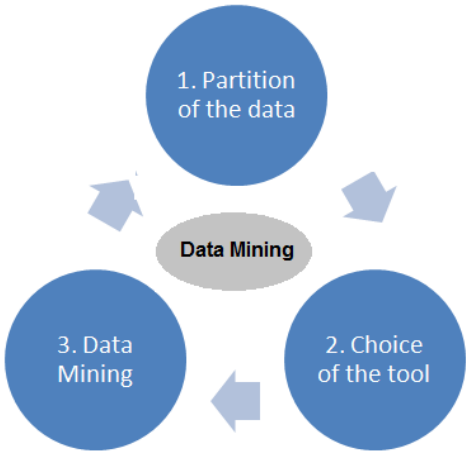

Fig.4. Data-mining activities

d) Post-Processing (Figure 6) is the last phase of the methodology. It includes the simplification and presentation activities of knowledge models achieved in the Data Mining phase. In this phase the results obtained are evaluated and are defined new alternatives of data surveys [8].

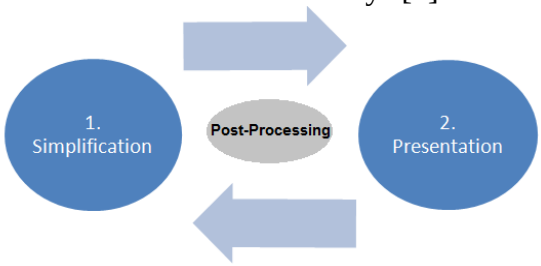

Fig.5. Post-processing activities

4. Data-Warehouse

The data warehouse represents a set with a huge volume of data (reaching up to terabytes) that is used as a data source that is compact, integrated and complete and it is a base for a variety of types of information systems (for example: decision support, executive, business intelligence) aiming to provide stored data from noteworthy sources (operational data, external files, archives, etc.) used at supporting the decision-making process in a company business [10].

Data warehouse represents a technology with an increasingly present in business being increasingly appreciated in the last years, becoming a performing solution in terms of clients and business development of a company. The fact that this technology is becoming more prevalent in more and more activity fields reported an efficiency of a large variety of operations and an improvement of market intelligence [11].

Data warehouses comprise data coming from many different information sources that are converted for a multidimensional representation used by Business Intelligence Systems [12].

Data warehouse represents a more complex form of database with a large volume of data, usually designed using the traditional relational model, which contains many historical data of a certain interest [4].

Data warehouse is organized as a unique source of data and information for the entire organization which represents a fundamental principle of the integrity, where data is stored in a single, common form of representation of data from all sources (databases, external files, archives, etc.) settling specifically conventions on the designation fields, coding systems, representation of measure units, representation mode for calendar data, avoiding duplication of the same thing originated by different sources (departments) [13].

Figure 6 illustrates a model for data-warehouse tables organisation for railway level crossing. Is observed the higher number of atributes which are the base for a detailed data storage.

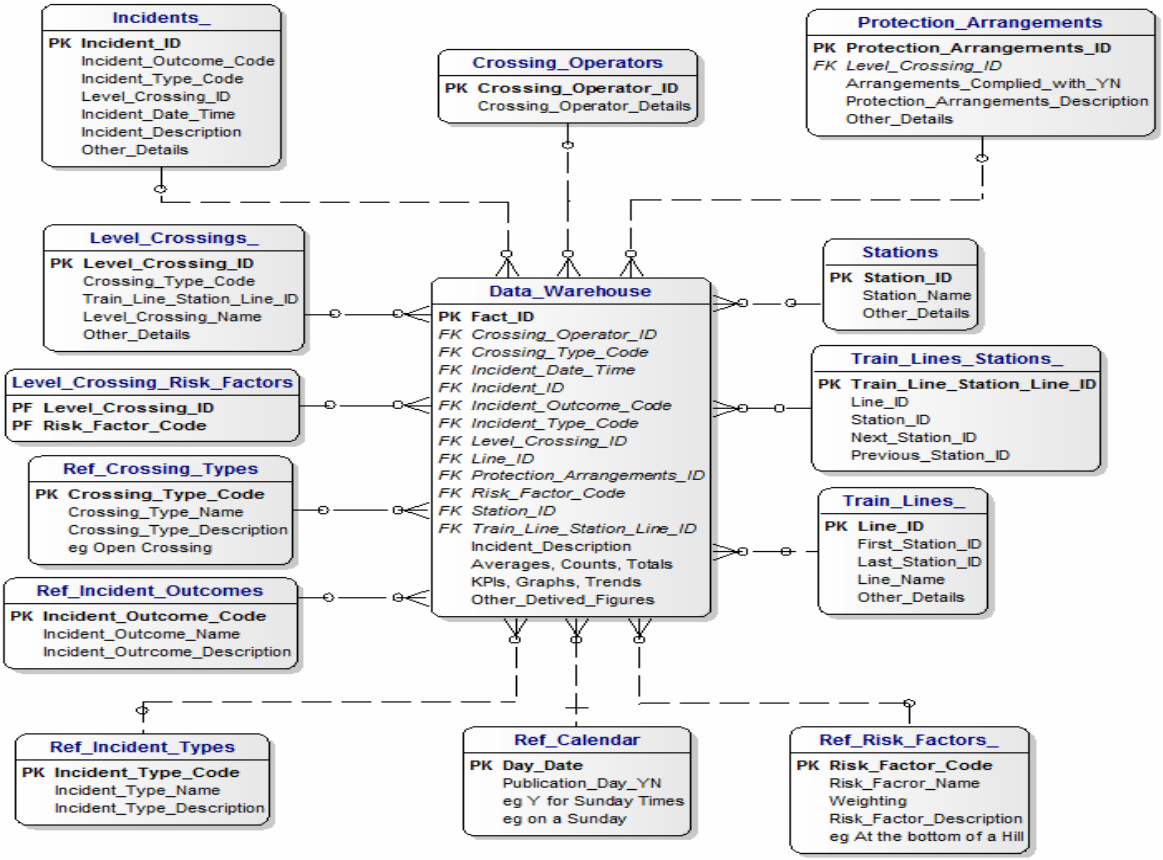

Fig. 6. Data Warehouse for Railway Level Crossings [14]

Conclusions

More and more European countries are currently favors for the transport of freight and passengers on the railroad for support environment. In this context, Romania is trying to adapt to this trend by developing regional operational programs. Business intelligence systems are the basis for adapting to this trend because it showed his usefulness in many developed or developing countries. The shift to the modal or mixed transport requires a large volume of records to observe how it can be constantly improved. It is also important to know how the data are used for the discovery of knowledge and supporting decision-making at the CFR. In conclusion we can notice that IT technologies (data-mining and data-warehouse) are successfully adapts to this area.

For the applicative part, there was presented a star schema of data-warehouse for railway level crossing where the data is stored in many tables having a very high level of detail. These data are used by business intelligence system to achievement reports, forecasts and support the management in making beneficial decisions for the company.

References

[1] Larissa T. Moss, Shaku Atre - *Business Intelligence Roadmap: The Complete Project Lifecycle for Decision-Support Applications*, 2003, ISBN 978-0201784206.

[2] Claudiu Brândaș - *Contribuții la conceperea, proiectarea și realizarea sistemelor suport de decizie*, Teză de doctorat Universitatea „Babeș-Bolyai”, Cluj-Napoca, 2007.

[3] http://www.comunitateaerp.ro/utile/28

[4] B. Nedelcu, “*Business Intelligence Systems*”, Database Systems Journal, Vol. IV, Issue 4/2013, pg. 12-20, 2013, ISSN 2069-3230. Available:

http://www.dbjournal.ro/archive/14/14_2.pdf
[5] http://www.globalfootprints.org/sustainability
[6] International Institute for Sustainable Development https://www.iisd.org/sd/
[7] Viorel Simuț - *Managementul transportului feroviar*, Editura Asab, București , 2001, ISBN 973-85247-0-9.
[8] http://www2.sas.com/proceedings/forum2007/161-2007.pdf
[9] Kimball, Ralph, Ross, Margy. *The Data Warehouse Toolkit. Guia Completo para Modelagem Dimensional*, Editora Campus. Rio de Janeiro, 2002.
[10] Popa Aida Maria – *Data Warehouse Pyramidal Schema Architecture – Support for Business Intelligence Systems*, The 14th International Conference on Informatics in Economy, 30 aprilie – 3 mai 2015, Proceedings if The 14Th International Conference of Informatics in Economy, Bucuresti, ISSN 2284-7472.
[11] John Foley, **"***The Top 10 Trends in Data Warehousing*", March 10, 2014. Available: http://www.forbes.com/sites/oracle/2014/03/10/the-top-10-trends-in-data-warehousing/
[12] G. Satyanarayana Reddy, M. Poorna Chander Rao, R. Srinivasu, S. Reddy Rikkula, "*Data Warehousing, Data Mining, OLAP and OLTP Technologies Are Essential Elements To Support Decision-Making Process in Industries*", International Journal on Computer Science and Engineering(IJCSE), vol. 2, No. 9, pp. 2865-2873, 2010, ISSN 2865-2873.
[13] M. Velicanu, Gh. Matei, *Tehnologia inteligenta afacerii*, Editura Ase, Colectia Informatica, Bucuresti, 2010, ISBN 978-606-505-311-3.
[14]http://www.databaseanswers.org/data_models/railway_level_crossings/data_warehouse.htm

Aida Maria POPA graduated from the Faculty of Cybernetics, Statistics and Economic Informatics of the Academy of Economic Studies in 2012. She graduated from the Economic Informatics Master of the Academy of Economic Studies in 2014. Currently, she is a PhD candidate, coordinated by Professor Manole VALICANU in the field of Economic Informatics at the Bucharest University of Economic Studies. Her scientific fields of interest include: Databases, Data Warehouses, Business Intelligence, Decision Support Systems and Data Mining.